Identification of Hotspots for Heatwaves using Big Data

Sang-Wook Kim, Taehyun Kim, Jongchul Park, and Yeora Chae
Korea Environment Institute

Introduction

- This study identifies hotspots for a heatwave using the high-resolute big data of Seoul, South Korea.
- Resident credit information and floating population data are used as socio-economic factors in 100 m grid resolution.
- High resolution (1 km) temperature forecast data downscaled from Dong-Nae Forecast (Digital Short-range Forecast) system is used as a meteorological factor.
- The hotspots are determined by high-temperature areas in the vulnerable regions.

Data & Method

- Period: 2017-2018 JJA (June-July-August)
- Domain: Seoul, South Korea (Figure 1)
- Data for Socio-economical Condition
  - Average annual income per person in grid point (Korean Credit Bureau)
  - Floating population: the number of people in grid point (KT, annual mean population of each year removed)
- Data for Meteorological Condition
  - Point observed temperature: Seoul station (Korean Meteorological Administration, climatology removed)
  - Temperature distribution: Forecast field downscaled (1 km) from KMA short-range forecast (Lee *et al.*, 2019)
- Hotspot identification algorithm (Figure 2)
  - Socio-economical condition (Static Vulnerable Regions + Dynamic Vulnerable Regions) + Meteorological condition (Heatwave Forecast)

Table 1. Data description

Name	Production	Spatial Resolution	Temporal Resolution	Description
Resident Credit Information	KCB	100 m (grid)	Fixed (2018.3.)	Resident population and average annual income
Floating Population	KT	100 m (grid)	Hourly (17-18JJA)	Floating population estimation data based on LTE usage
Station Temperature	KMA	Point	Hourly (17-18JJA)	Station observed temperature (47108)
Temperature Forecast Field	KMA & HUFS	1 km (grid)	3 hourly (17-18JJA)	High resolute temperature forecast field downscaled from Dong-Nae Forecast (KMA short-range forecast)

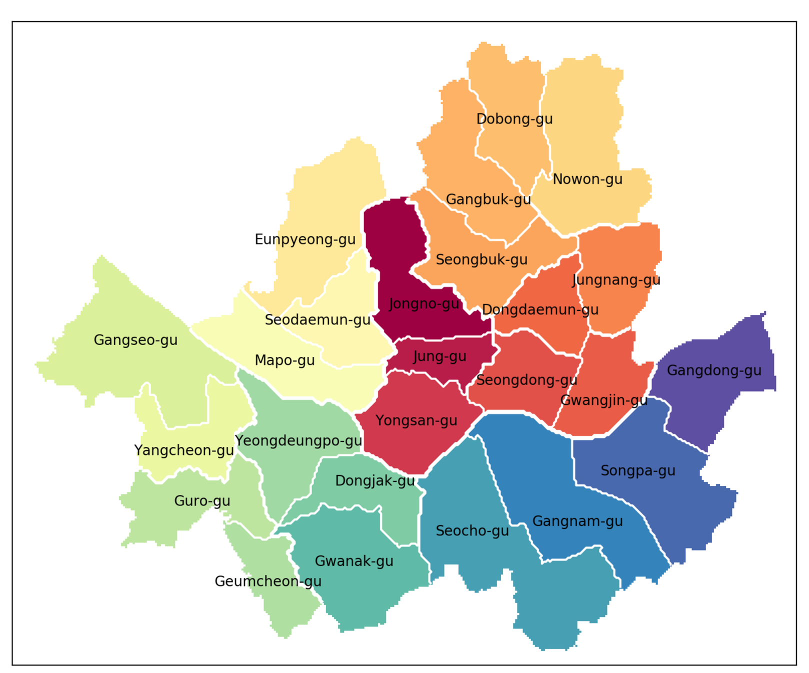

Figure 1. Seoul domain and administrative divisions

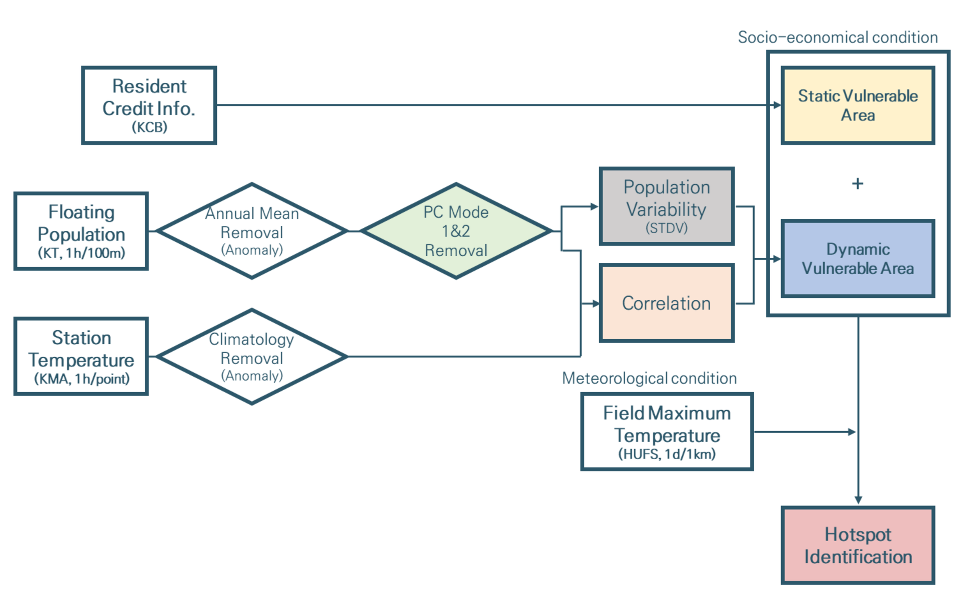

Figure 2. Hotspot identification algorithm including data used and pro-process

Results

- Static Vulnerable Region
  - Low-income residential regions (lover 20%)
  - The region showing lower average income of the population living in each grid → the more vulnerable to heatwaves expected
  - Fixed in time within a year and a season (static in time)
  - Mainly distributed in the southwest and the northeast area of Seoul (red & blue boxes in Figure 3b)

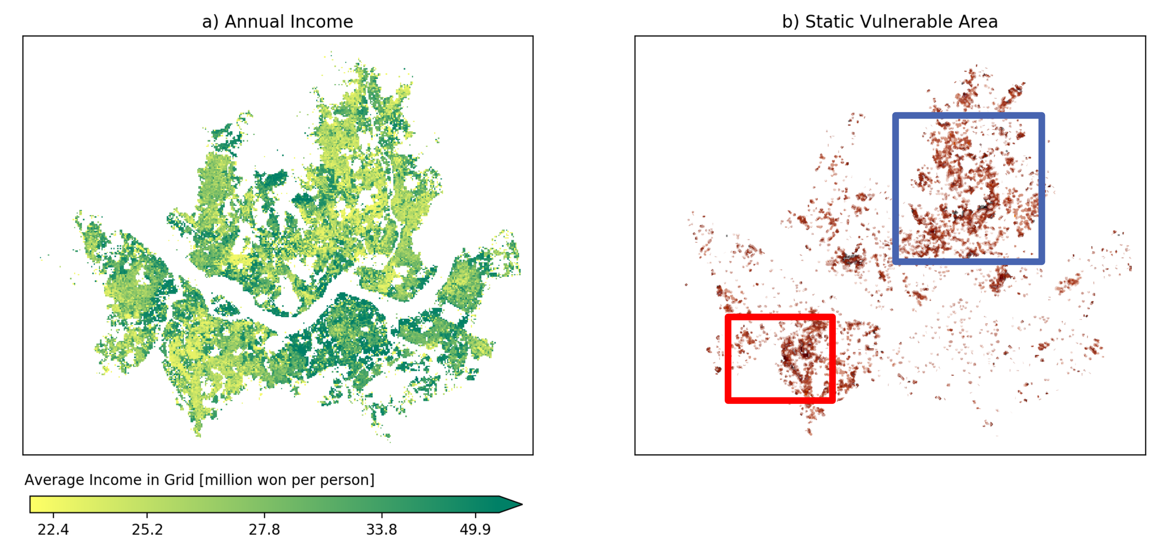

Figure 3. Individual average annual income at each grid (a) and the static vulnerable regions (b) in Seoul domain (1 million won ≈ 820 USA dollar).

- Dynamic Vulnerable Region
  - Small population variability & weak correlation with temperature
  - The region where population mobility is less responsive to temperature
  - Various in time (dynamic)
  - Floating population data pre-processing
    - · Population density by grid over time (personal movement path does not appear)
    - · Pre-process for removing temperature-independent population variability
      → Eliminate the PC(EOF) mode 1 & 2 (both variabilities indicating commuting pathway) of the floating population time series from the original population change (Residual population (RP), Figure 4 & 5)

    ※ mode 1: population moves from residential area to work place
    mode 2: population moves from work place to residential area or nightlife area

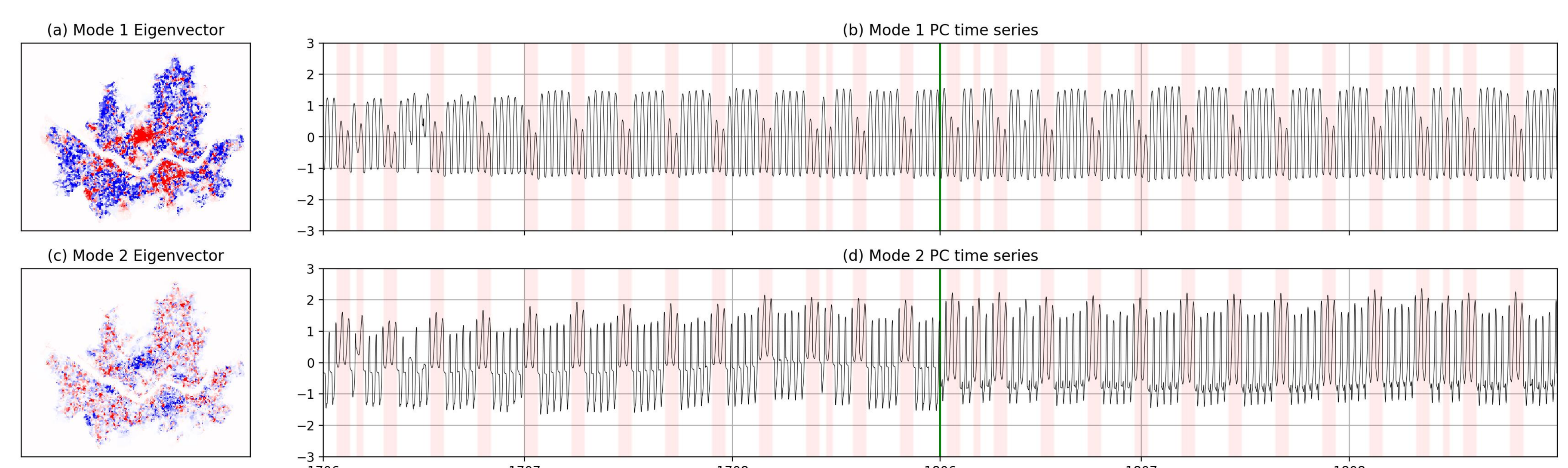

Figure 4. Eigenvector and PC time series of PC mode 1 (top) and 2 (bottom) of the floating population

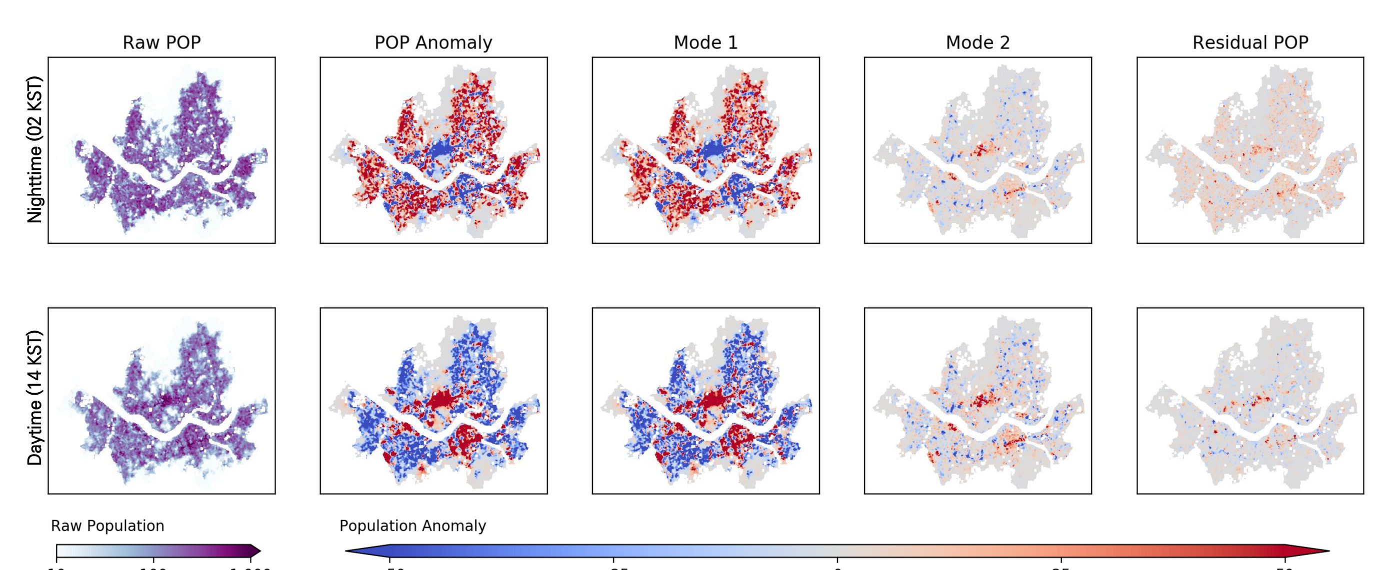

Figure 5. Pre-process of floating population data at 02KST and 14KST. From the left column, the original population, population anomaly (annual mean removed), PC mode 1, PC mode 2, and residual population (mode 1 & 2 removed from population anomaly).

- Dynamic Vulnerable Region (continue)
  - The region that the residual population (from the previous step) has small variability and weakly correlate to the point observed temperature at each time
  - (Dynamic) Vulnerability

$$Vulnerability = \frac{1}{STDV_{RP} \times Corr}$$

(where, $STDV_{PR}$: standard deviation of residual population and $Corr$: correlation between residual population and temperature)

  - Dynamic vulnerable condition: $STDV_{RP} < 30, Corr\ in\ (-0.2, 0.2), and\ Vulnerability > 1$
  - Less responsive to temperature changes (people in dynamic vulnerable regions are expected to stay in there at reference time regardless of temperature changes)
  - Mainly distributed in the outskirt of Seoul

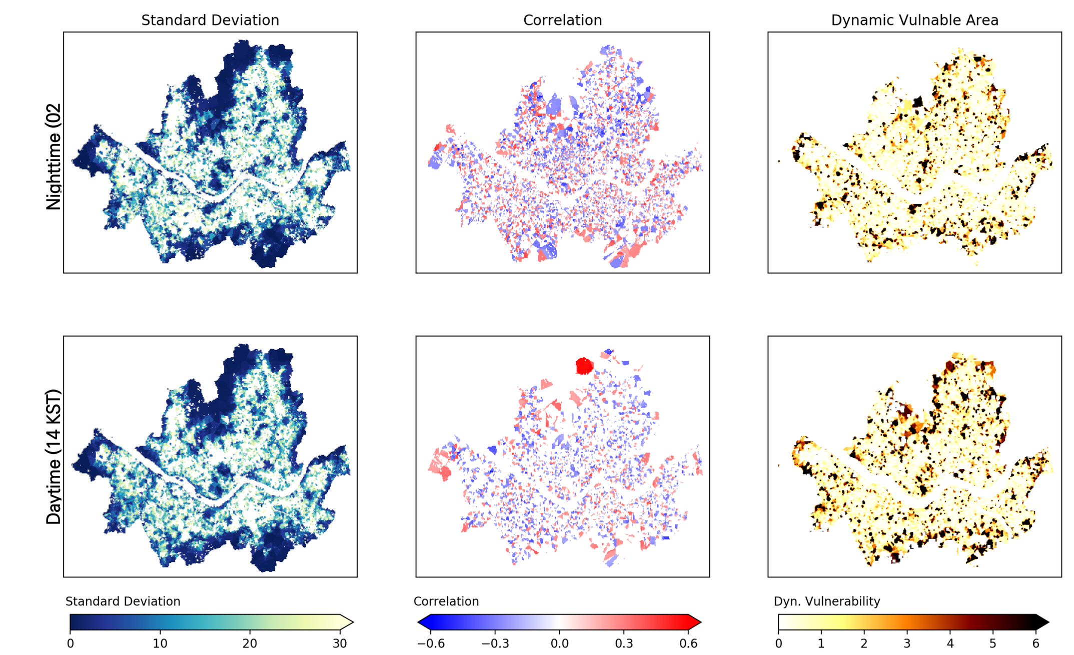

Figure 6. From the left column, standard deviation of residual population, correlation between residual population and temperature, and vulnerability at 02KST and 14KST.

- Hotspots for heatwaves
  - The vulnerable regions where heatwaves are expected
  - The region where 33℃ (heatwave threshold of KMA) or higher temperature is forecasted while satisfying both static and dynamic vulnerable condition
  - Distribution changing in time and temperature; e.g. 2018-08-01 (forecasted at 2018-07-31 05KST, Figure 7)
  - Hotspot clusters: often appearing in densely populated residential areas with alleys and low building height

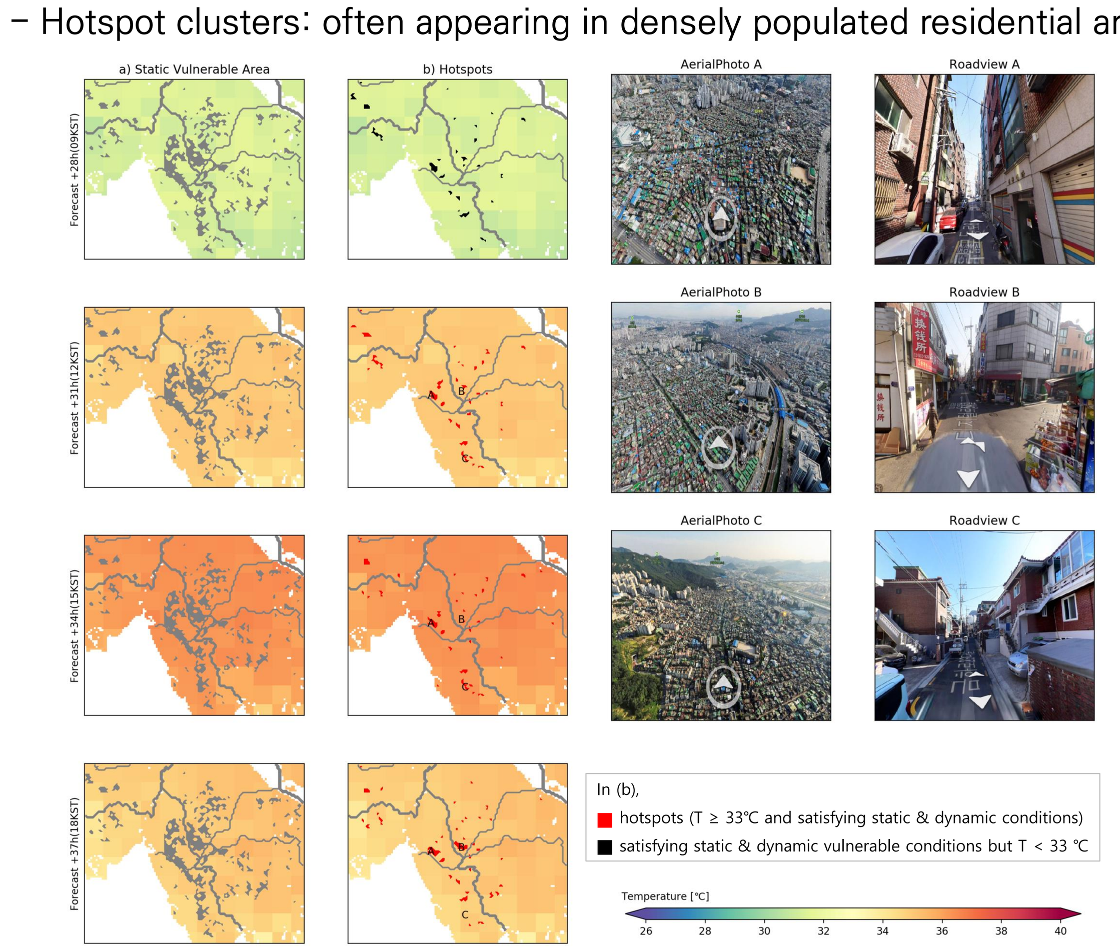

Figure 7. Static vulnerable regions (low-income residential area, (a)) and hotspots (b) distribution within the red box in Figure 3b. Shading represents the 09KST, 12KST, 15KST, 18KST (from top to bottom) temperature forecast the day before. The 3rd and 4th columns show aerial photo (3rd) and road view (4th) of points A, B, and C which are frequently identified as hotspots.

Discussions

- This hotspot identification system figures out where damage expected in objective way
- The higher frequency of hotspot appearance, the number of thermal illness patients robustly increases with temperature (not shown).
- Hotspot information is expected to support heatwave management such as making decisions about the place and timing to provide disaster relief resources.

Acknowledgement
This work was funded by the Korea Meteorological Administration Research and Development Program under Grant KMI (KMI2018-01410).